Multisensorial interactions

Annie Luciani

Advances in natural, human-centred, and personalised interaction system
Interactive audio-visual presentation and communication

MULTISENSORIAL INTERACTIONS

Annie LUCIANI, Acroe-ICA
IST - Bruxelles - 26 mai 2000

I would like to focuse my contribution to this workshop on two topics which seems to me key-topics for Interactive Audio-Visual Media :
• the concept of multisensorial interaction and the specific place of the gestural data in the interaction
• the concept of Dynamic Expressiveness in the interaction process and in the sensorial audio and visual media produced during an interactive process.

1. Human-centered expressive interaction

The main sensori-motor channels that human beings have at their disposal to produce and receive sensorial information are :
- sensorial channels : which are mainly : Acoustical, visual and gestural. The sensorial gestural channel is which that combines the proprio-tactilo-kinesthetic perceptions
- effector channels to emit information : We have only two effector channels to emit information to the external univers : our gestural actions (motricity, muscles ...) and the voice.

Because voice communication has been already considered in a lot of works (speech synthesis, speech recognition ...), I don't consider voice in this talk.

So, if we except voice, I would like put the focus on the idea that the "gestural channel", understood as all our means of motricity (hands, body, muscles ...) is the only way for us to emit something to the external universe.

This external universe can be composed of :
- human beings (oneself or anothers)
- the physical world which can be natural world or artificial world as machine human beings have built.

Whatever the "vis à vis" is, and whatever the information you have to transmit to him or it, without the voice and the speech, we van notice that :
- we must use our hands or our body to create or to modify visual, acoustical or gestural informations in the aim to communicate with the external universe.
- and, because human beings cannot create directly all the visual and acoustical information they can perceive (only in a very restrictive way : the image of our own body or the sound of our own body!!!) : we need the use of an intermediate physical object to transform the physical informations produced by your hand in acoustical, or visual information able to be received by our eyes and our ears.

Briefly saying, The fundamentals of human interaction are :
1. hands are the only way (except the voice) to emit information.
2. An intermedaite object (called in music "instrument") is allways necessary to adapt the mechanical information produced by our motricity to the visual and acoustical information perceived by our sensorial channels. The Instrument is the way for the body to speak to our eyes and our ears.
3. Acoustical and visual channels are only sensorial channels.
4. Gestural channel is simultaneously a sensorial channel and an emmisive channel. The gestural channel is the only sensori-motor channel.
5. Person centered interaction are multi-sensori-motor closed **loops** :
    - gestural action / gestural perception loop
    - gestural action / visual perception loop
    - gestural action / acoustical perception loop

Some neurophysiological and psychological properties of these closed loops
1. The Gestural action / Gestural perception loop runs more faster than the two others. It is because this loop is a reflexe loop which circulates not by the brain, as the two others, but by the spinal chord. The rate of this loop is about 1 ms of delay maximum between the input and the output for 10 ms (10 times more) for the two others loops.

2. Our motor system have about 200 degrees of freedom, with about 15 DOF for each hand.
The gestural action channel differs than the visual or acoustical channel because the organization of the effectors chain can be changed according to the task.
For example, the effector chain composed of the two hands and two arms is organized as a parallel structure when we assemble two equivalent parts of an objet and it has a serial structure when we perform a accurate task with one hand (for example when you are writing) and we maintain the context of the task with the other hand (for positionning the paper sheets).

II. Gestural modality (Haptic Modality) to produce and communicate sensorial data

The gestural modality (often called haptic modality) is in my opinion a new key-component of centered person-system interaction.

Because the gestural channel is a sensori-moto channel, it performs three functions ;

• We can use the gestural channel only as an effector :

-> that is the semiotic function of the gesture

It is the functionnality mainly developped in computerized environments, through mouses or keyboards. In this kind of action, the content of the gesture is not conveyed in the final performed task. When you point the banana, chimpanzees look at the finger but human beings the banana.

• gestural channel is also a way to know : inertia, rigidity, the weight and inertia, resistance to the displacement, mechanical material or structure of objects.

-> and that is the epistemic function of the gesture

It can be use such as an Haptic Display. This function is not widely developped with computers.

• But it is mainly an Action/perception channel to perform tasks in which the action and the perception are completely linked, and in which energy is conveyed from person to external universe and from external universe to the person

-> that is the ergotic function of the gesture.

In this situation, the dynamic content of the interaction is conveyed into the controlled or created visual or audio media. A highly personnalized information is engraved in the others sensorial data. When we hear violin, it is quite difficult to know if the instrument reveales the investment of the musician or if the musical investment reveales the quality of the instrument. This kind of interaction is the favourated way to support expressiveness : Expressiveness during the interaction process, Expressiveness in the performed tasks

Expressiveness is THE contemporary challenge in a computerized and networked environements:

- for computer productions and communication of sensorial media of communications : images, animation, sounds, musics
- for increasing the natural and the efficiency of human-human or human-machine communication to perform non sensorial tasks : Computer Aided Design, Computer Aided Decision, Process Control...

From these analysis, we can extract some new topics to developp today in the aim:

- to enhance the expressiveness in the interaction processes
- To developp new tools to work with, to produce and to convey expressiveness

1. Haptic interfaces (or Gestural transducers)

Developpement of forces feedback transducers : for professionnal uses or general uses. The technological breakthroughes to cross over are :

- the compacity : table devices or embedding systems
- Time response and frequency bandwith : it is a crucial parameter for a usable force feedback.
- the number of degrees of freedom
- the versatily of the morphology

All these requirpements are greatly depending on the task to be performed and they require new developpements in task analysis and ergonomy in the contexte of programmable force feedback.

2. Computer systems and models

Hard synchronisation between action and perception

Good quality of the tranformation gestures to images and gestures to sounds.

New software developpemts can be opened:

- real time with hard synchronisation at the lower level of the machine
- new real-time operating systems in which the multisensorial action-reaction principle would be plugged inside the operating system itself
- new dynamical and reactive generative models to generate and control visual motion and acoustical evolution under gestural playing
- new generative models to convey sufficient complexity for human perceptions and controls

3. Gestural-like signal processing

Besides, visual and acoustical signals, Gestural signals, acquired by motion sensors, is a new kind of signals to process and to transmit, which can be called "mechanical signals"

Gestural signals are the signals which describe the cinematic evolution of the body motion. It is of the same nature of all the motion signals from mecanics or biomecanics. it is the same kind of signals as the motion in a visual sequence or continuous macroscopic evolution of sounds.

As shown here, It is at an intermediate scale between the two others:

So, the technological and psychological developpements to perform are : Gesture and Motion Capture, Gesture and Motion Analysis, Gesture Synthesis, Gesture and Motion Coding and Representation, And the developpement of Gestural Langages from pertinent Motor Schemes Extraction.

4. Person-system Interfaces designing

In spite of the concept of "realism on each sensorial channel : realism in images, realism in sound ..." tha aim becomes :
How to place human beings in natural and efficcient situation of working?
How to convince that the virtual system they manipulate, is like the reality during working, or perceving ...?
The problem is no longer to try to reproduce the reality but to convince that the worker, or the listener, or the viewer is in conditions that are similar to the reality.

From multisensorial interfaces, the aim is to combine all the main sensori-motor channels to perform the best conditions of acting on, of learning, of perceving ... and so on ... -> that is intersensoriality

So the concept of the material immersion in which the user is immersed in virtual universes by materials (head mounted screens, data gloves ...) becomes replaced by the concept of "cognitive immersion", in which the user re-create a robust cognitive real-like object.

5. Transmit gestural-like signals to other persons in other places, at other times : A challenge for the networks operators

As any acoustical or visual signals, the gestural-like signals can be used not only in a local environments which are able to implement an hard action-reaction situation.
They can be transmitted in other places, in other times, for other persons. This will be the third kind of phenomenological signal to be transmitted.
By it rise a beautiful challenge :
- because it is multidimensionnal and flexible in its DOF structure
- because it requires closed looped between emission an reception
- because it requires to know two dual variables for its distant reconstruction,

It rises a beautifuf challenge at telecommunication technologies :
• defining steps of loop rates more or less local according to hierarchical extraction processes of low-rate features as shown here : to transmit gestures as far away as possible, we must decrease the time delay of synchronisation and then we must define extraction processes of low rate informations and and their dual reconstruction processes for the other extremity of the channel
• increasing the quality of the synchronisation
• defining standard for gestural data taking into account the bidirectionnality of these data

6. From gestures to Images - deep changes in computer animation tools

The architecture of Computer Animation Tools are changing fundamentally under the novelty of the motion capture.
Previouly, the core of the systems was composed of the 3D shape modeller and provided a low level;of interactivity in the design of the motion.

The main problems to day are (1) to decrease the cost of the animation which is the main cost in the animated pictures production, (2) to decrease to delay of production, (3) and simultaneously to increase the quality of the animation. So, the core of the Computer animation tools must become the motion synthesis stage itself in spite of the 3D modeler.

The trends are in the implementation of an high level of interactivity and of real time generation and control in the designing the animation. So animation is not applied after the shape, but it trends to be defines above the the 3D final shape. And the animation control trends to be gestural and motion control with realt time actoring control

<u>Lines to developp</u>
Motion body (hand) capture
Motion body (hand) analysis : how to transform multidimensional and correlated sampled signals in pertinent dynamical features and events able to be manipulated by the motion designer: gestural processing
Actorisation techniques
new sensors and new transducers, force feedback transducers adapted to the motion control in complex 3D scenes : lips motion, walking, virtuals avatars, complex environnements (deformables, breaking, chaotic rigid objects, heterogeneous scenes)
Multi-animator control
New relation ship between shape modelers and motion modelers.
- Possibility of motion scenarii re-designing, scenarii designing on lines
- Re-usability with a wide capability of variation : the concept of theme and variation in animation

7. DMI

Seeking the expressivness

In spite of an higher level of the research, the european industry of DMI has been ratiboised (put down, mow down, clearly cut down) by the agressivity of the american and japonese industry in 1983. The evolution of the concepts required by the general public was :

- hearing non natural sound i.e. like - synthetic soung in the 70ies and the eighties
- come back to the real reference in nineties withn the developement of sampling, but with low interactivity
- increase the human control and the interactivity and simultaneously increase the human quality and complexity of the sounds ; that is the current demand fir thes next years.

Lines to developp :

dynamical modeling, real-time production, expressivness analysis and synthesis, gestural control, gestural anbalaysis of musical gestures, cognition in the creative process of sounds ... have access to music during the production process of sound, motors schemes, new organizations of instruments and of gestures

-

8. Cooperation Sound and images - new methods and new problems

9. Psychological trends

The ecological theory of perception assumes that the perception cannot work without action.

Strong interactive working and daily environnements allows psychologist to evaluate an ecological theory of sensori-motor interaction.

Uses

- Increase the quality of Virtual Reality plateforms
- Increase the quality of tools for producing Computer animation and computer sound and music
- Learning technical gestures : high level gestures (surgeon, instrumentist, others complex manulatory tasks : inserting components, driving machines in complex or dangerous environements ...
- learning manual daily tasks : writing, ...
- fill the gap in the educationnal costs between manuals workers and dexterous experts by adequate education, create new
- fill the gap between body virtuosity and body deficiency
- design adapted new ergonomy of the device : Keyboard changing by using the analysis of motor schemes

Examples : keyboard with a few number of key to play complex music is possible after the separation between the organization of the music (the composition) and the motor scheme needed to play it.

Some problems increased by the introduction of a hiogn degree of personnalization and of human expressiveness in flexible and versatile computerized and neworked technologies